

Topic Modeling of Grant Applications

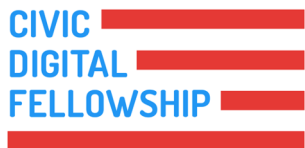
NINR Division of Extramural Science Programs

Rebekah Rasooly, PhD

Martha Matocha, PhD

Kristopher Bough, PhD

Wynn Meyer, PhD



Mina Narayanan
Carnegie Mellon University
Public Policy and Management

Background

- The National Institute of Nursing Research (NINR) aims to :
 - (1) Build the scientific foundation for clinical practice
 - (2) Prevent disease and disability
 - (3) Manage and eliminate symptoms caused by illness
 - (4) Enhance end-of-life and palliative care
- Overlap of research areas between institutes
- Desire to attain a more granular understanding of research areas and portfolio configuration

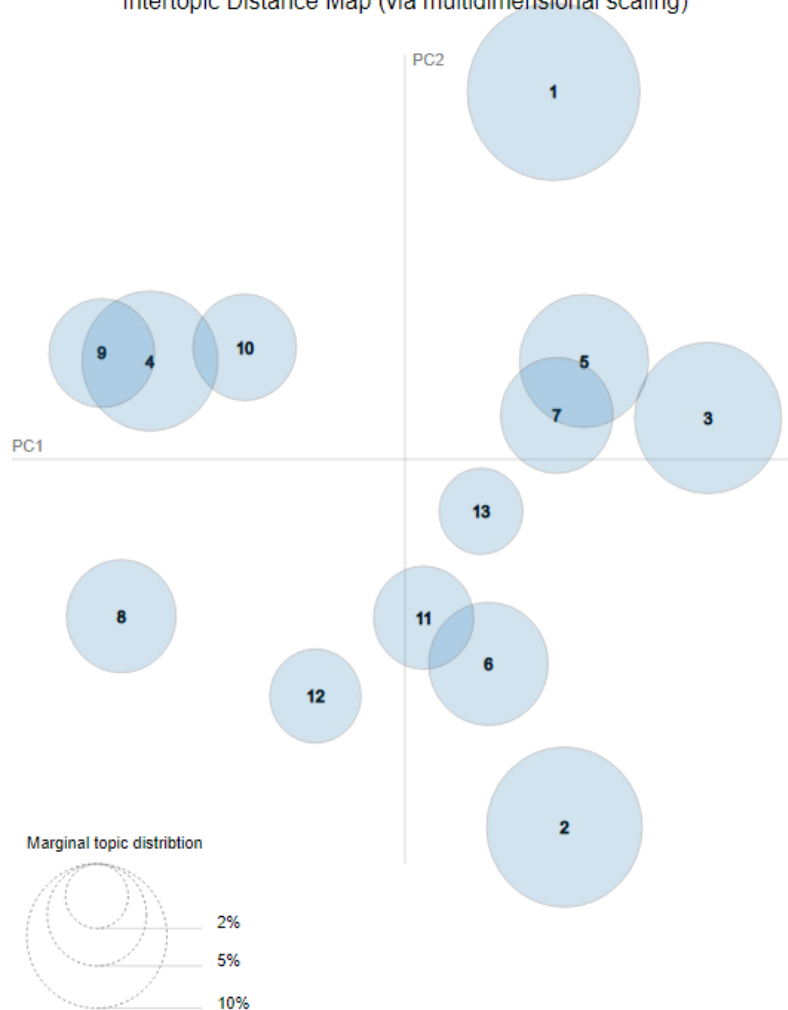
Research Questions

- What are the topics of NINR grants and how have they changed over time?
- How do the topics of NINR grants compare to those of other ICs?

Topic Modeling and LDA

- Topic modeling is a machine learning technique that discerns topics from a collection of text data
- Latent Dirichlet Allocation (LDA) is a variant of topic modeling. Specifically, it is a generative statistical model that creates topic groups from text data, where each topic group contains words that are predictive of that topic or closely associated with that topic
- Unsupervised approach

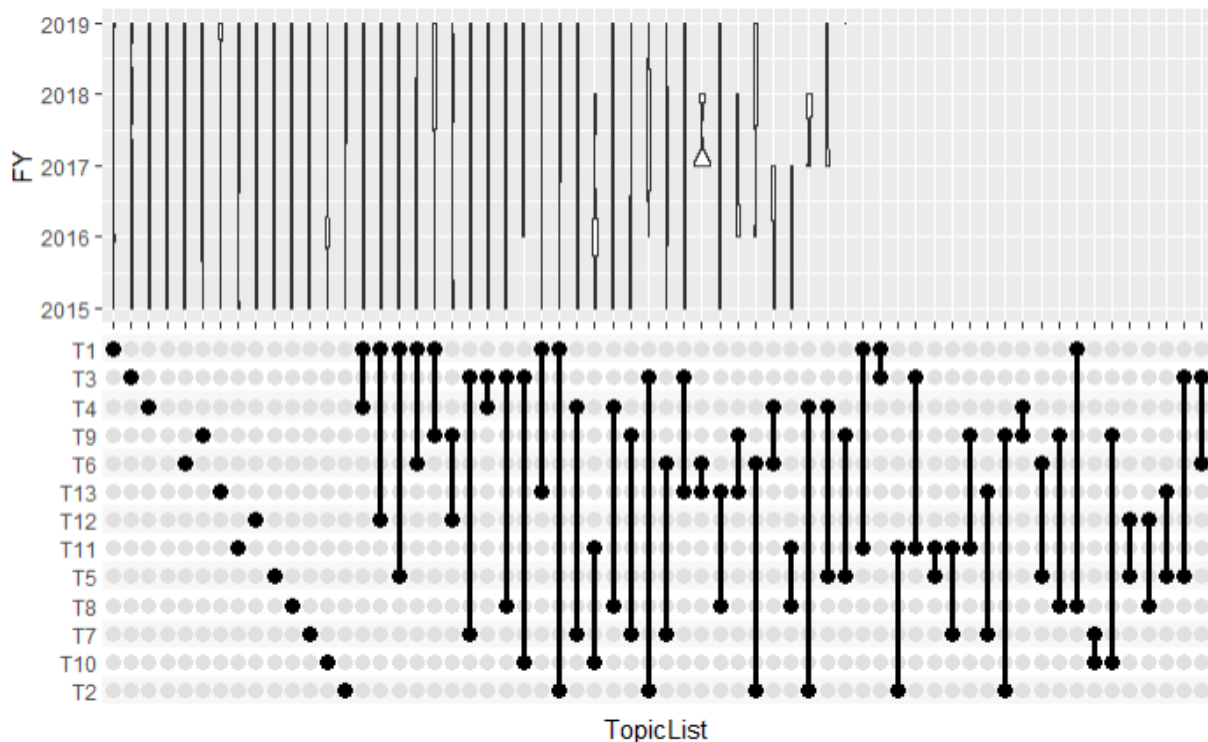
Intertopic Distance Map (via multidimensional scaling)



T1 = algorithm/sensor/software
 T2 = food/vaccine/school
 T3 = mHealth/diabetes/app
 T4 = cognitive/stroke/brain
 T5 = caregiver/palliative care/dementia
 T6 = heart failure/sodium/black men
 T7 = advanced care planning/end of life/multiple chronic conditions
 T8 = women/maternal/pregnancy
 T9 = wound/heal/gene expression
 T10 = pain/sleep/insomnia
 T11 = parent/asthma/children
 T12 = physical activity/cardiovascular disease/sedentary
 T13 = HIV/viral

Dataset = NINR grants from 2015 to 2019
 13 topics

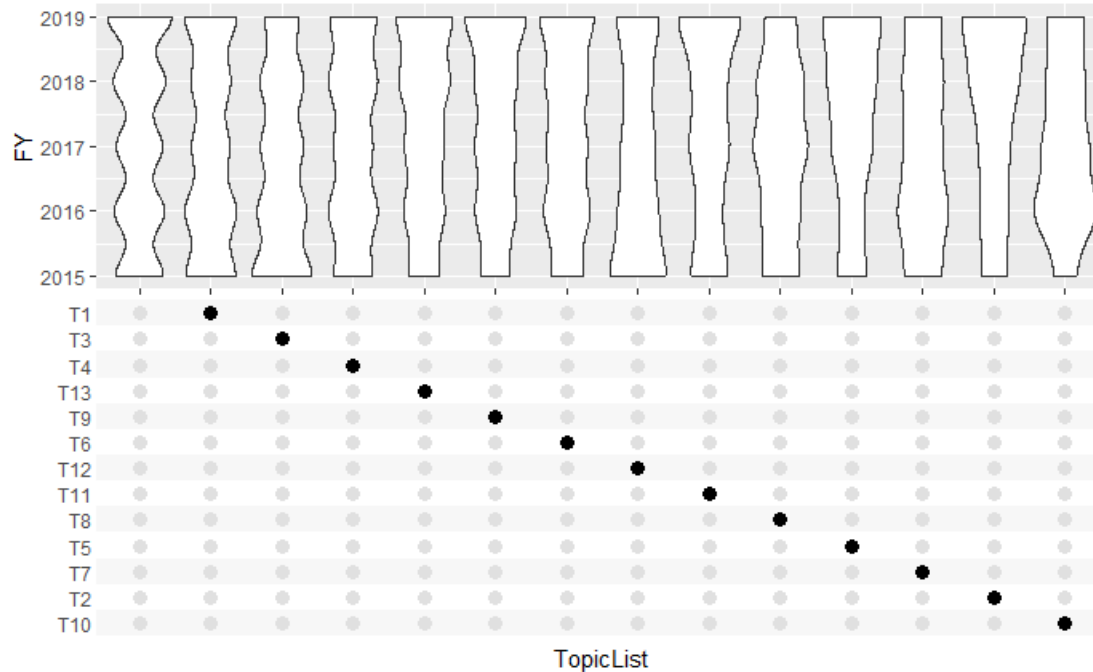
Intersections of Topics from 2015 to 2019



T1 = algorithm/sensor/software
T2 = food/vaccine/school
T3 = mHealth/diabetes/app
T4 = cognitive/stroke/brain
T5 = caregiver/palliative
care/dementia
T6 = heart failure/sodium/black
men
T7 = advanced care planning/end
of life/multiple chronic conditions
T8 = women/maternal/pregnancy
T9 = wound/heal/gene expression
T10 = pain/sleep/insomnia
T11 = parent/asthma/children
T12 = physical
activity/cardiovascular
disease/sedentary
T13 = HIV/viral

Dataset = NINR grants from 2015 to 2019
Threshold = 40%
13 topics

Growth/Contraction of Individual Topics from 2015 to 2019



T1 = algorithm/sensor/software
 T2 = food/vaccine/school
 T3 = mHealth/diabetes/app
 T4 = cognitive/stroke/brain
 T5 = caregiver/palliative
 care/dementia
 T6 = heart failure/sodium/black
 men
 T7 = advanced care planning/end
 of life/multiple chronic conditions
 T8 = women/maternal/pregnancy
 T9 = wound/heal/gene expression
 T10 = pain/sleep/insomnia
 T11 = parent/asthma/children
 T12 = physical
 activity/cardiovascular
 disease/sedentary
 T13 = HIV/viral

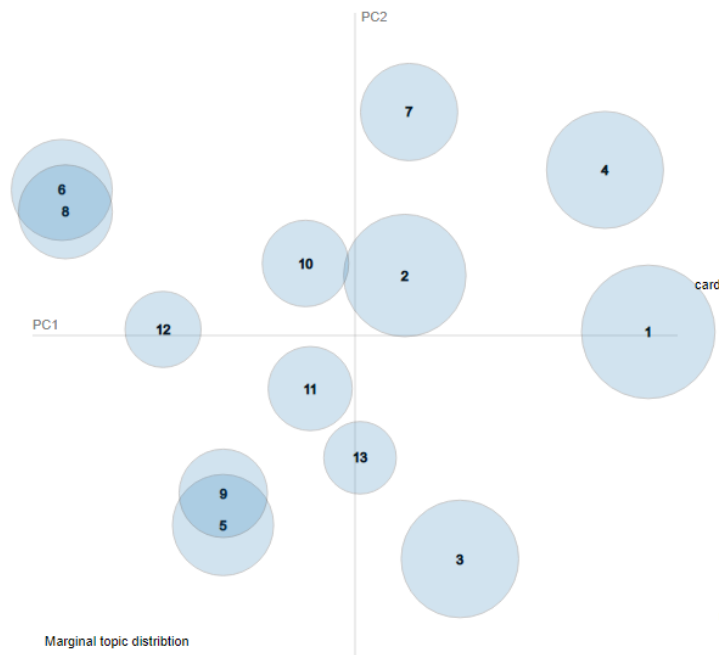
Dataset = NINR grants from 2015 to 2019
 Threshold = 51%
 13 topics

Institute Topic Comparison – Prevention

- Compare the Prevention grants funded by NINR in 2019 to those funded by other ICs

Institute Topic Comparison – NINR and NHLBI

Intertopic Distance Map (via multidimensional scaling)



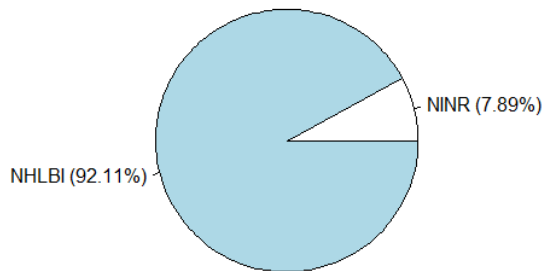
Marginal topic distribution



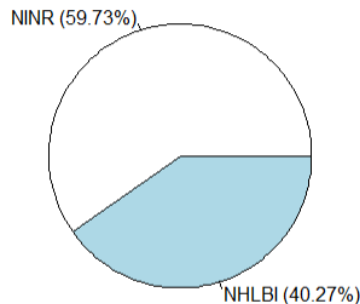
Dataset = NINR and NHLBI grants from 2019
13 topics

Institute Topic Comparison – NINR and NHLBI

T10

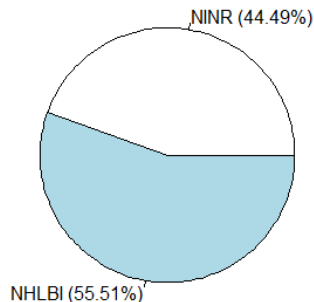


T12



T6 = vascular/stroke/biomarker
T10 = lung/pulmonary
T12 = obesity/metabolism/stress

T6



Topic Modeling Application

Topic Modeling NINR

Create topic models with grant application data.

Grant application file input

Browse...

No file selected

Select whether intermediate output should be written to Excel files.

I want intermediate output written to a file.

- ☒ Yes
☐ No

Topic Modeling NINR

Create topic models with grant application data.

Grant application file input

Browse...

NINR-2019-all-apps-with-RCD

Upload complete

Select whether intermediate output should be written to Excel files.

I want intermediate output written to a file.

- ☒ Yes
☐ No

Stopwords

I want to load custom stopwords.

- ☒ Yes
☐ No

Continue

Potential

- Inform the next Strategic Plan
- Determine whether NINR portfolios need to be reconfigured so that they align with applicant interests
- Validate hands-on approach for Program Officer assignments
- More efficiently classify grant applications into research categories